

EL PRINCIPIO DE PRESUNCIÓN DE FALIBILIDAD DE LAS DECISIONES ALGORÍTMICAS DESFAVORABLES UNA NUEVA GARANTÍA JURÍDICA FRENTE A LAS DECISIONES AUTOMATIZADAS Y EL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACIÓN PÚBLICA(*)

JESÚS A. TAHIRÍ MORENO
Doctorando de Derecho Administrativo
Universidad de Extremadura

SUMARIO: I. LA INTELIGENCIA ARTIFICIAL COMO TECNOLOGÍA ESTRATÉGICA AL SERVICIO DEL CIUDADANO: 1. Introducción al concepto de aprendizaje automático supervisado y no supervisado. 2. La conveniencia de la implementación de sistemas de aprendizaje automático supervisados en actuaciones que impliquen únicamente el ejercicio de potestades regladas. 3. La problemática del uso de sistemas de inteligencia artificial que utilicen técnicas de aprendizaje automático no supervisado por la Administración. ¿Prohibición absoluta o relativa?– II. EL PRINCIPIO DE PRESUNCIÓN DE FALIBILIDAD DE LAS DECISIONES DESFAVORABLES DICTADAS POR SISTEMAS ALGORÍTMICOS: 1. Su conexión con en el derecho a una buena administración y la Carta de Derechos Digitales del Gobierno de España. 2. La teoría de la proporcionalidad inversa aplicada a la actuación administrativa automatizada: la prohibición de actos desfavorables cuyo único fundamento sea el resultado del algoritmo y su correlativa permisibilidad en el supuesto de actos favorables. 3. El programa BOSCO como ejemplo paradigmático de la necesidad de regular la presunción de falibilidad de las decisiones algorítmicas.– III. REFLEXIONES FINALES: SOBRE LA NECESIDAD DE REMODELAR UNA ARQUITECTURA ADMINISTRATIVA ANQUILOSADA.– IV. BIBLIOGRAFÍA.

RESUMEN: El presente artículo tiene por objeto el planteamiento de una nueva garantía dirigida a reforzar la protección jurídica del administrado en procedimientos administrativos donde existan actuaciones administrativas automatizadas y/o se empleen sistemas de inteligencia artificial para la toma de decisiones. En la primera parte del trabajo se exponen los conceptos básicos del aprendizaje automático supervisado y no supervisado, analizándose, asimismo, los óbices jurídicos que impiden su uso en determinadas parcelas del poder administrativo. A continuación, se desarrolla el principio de presunción de falibilidad de las decisiones algorítmicas desfavorables como técnica de

(*) Trabajo recibido en esta REVISTA con fecha 25/04/2023 y evaluado favorablemente para su publicación el 22/05/2023.

supervisión humana en orden a facilitar la implementación de tecnologías emergentes en la Administración, a la vez que se respetan los derechos e intereses legítimos de los ciudadanos. Finalmente se realiza una breve reflexión acerca de la importancia de la labor doctrinal en este campo, incidiendo en la necesidad de remodelar la arquitectura de una Administración anquilosada.

Palabras clave: presunción de falibilidad; inteligencia artificial; decisiones administrativas automatizadas; aprendizaje automático; garantías del ciudadano; derecho administrativo.

ABSTRACT: The goal of this paper is to propose a new guarantee aimed at reinforcing the legal protection of the administrative proceedings where there are automated administrative actions and/or artificial intelligence systems are used for decision making. The first part of the article presents the basic concepts of supervised and unsupervised machine learning, analyzing the legal obstacles that prevent its use in certain areas of administrative power. Next, it develops the principle of presumption of fallibility of unfavorable algorithmic decisions as a technique of human supervision in order to facilitate the implementation of emerging technologies in the administration, while respecting the legitimate rights and interests of citizens. Finally, a brief reflection is made on the importance of doctrinal work in this field, with particular emphasis on the need to remodel the architecture of a stagnant Administration.

Keywords: presumption of fallibility; artificial intelligence; automated administrative decisions; machine learning; citizen's guarantees; administrative law.

I. LA INTELIGENCIA ARTIFICIAL COMO TECNOLOGÍA ESTRATÉGICA AL SERVICIO DEL CIUDADANO

Ríos de tinta corren en el ámbito académico sobre los riesgos, desafíos y oportunidades derivados del uso de la inteligencia artificial (en adelante, IA). Los artículos y conferencias hablando del tema se cuentan por cientos. Sin embargo, todavía no existe un *corpus* doctrinal consolidado que responda a la cuestión de cómo ha de afrontar la Administración el advenimiento de una tecnología que ya está transformando la sociedad.

¿Cómo adaptamos el ordenamiento jurídico-administrativo al uso de las nuevas tecnologías? ¿Acaso pretendemos encontrar soluciones sin reformas normativas que impliquen cambios estructurales en los procedimientos de actuación? Y en la hipótesis de que exista voluntad legislativa para regular esos cambios, ¿nos bastan los principios que tenemos, o hemos de dar un paso más y construir nuevos conceptos? El impacto de las nuevas tecnologías requiere un sesudo análisis jurídico que implicará años de esfuerzo, de ahí que las modas no sean buenas compañeras en el camino hacia la materialización (real y efectiva) de la transformación digital de las Administraciones Públicas. La Administración digital necesita nuevos principios que la rijan y el Estado

tiene el deber de adaptarse y adecuar el Derecho a las realidades sociales que vayan surgiendo.

Dicho esto, al jurista, y en especial al administrativista, se le presenta una tarea hercúlea, pues en breve tendrá que afrontar la regulación europea de la IA [la previsión es que la Propuesta de Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial (*Artificial Intelligence Act* por sus siglas en inglés, en adelante AAI) entre en vigor en 2024] (1). Ante este escenario es importante tomar distancia y preguntarnos cómo vamos a abordar la cuestión, ¿desde la rigidez del timorato que no quiere asumir riesgos o desde la flexibilidad y la confianza en el progreso tecnológico?

En la doctrina española existen varias posturas en torno al uso de los algoritmos por las Administraciones Públicas. Por un lado, están quienes defienden que algunos algoritmos son reglamentos y deben someterse a los mismos requisitos que estas normas jurídicas (2); por otro, autores que defienden lo contrario (3). Dentro de estos dos grandes grupos, algunos abogan por excluir la aplicación de la IA en determinados supuestos, como en la toma de decisiones discrecionales (4); otros, reconocen la posibilidad de su uso en actuaciones que requieran un «bajo nivel de discrecionalidad o cuando el ejercicio de la potestad discrecional supone el uso de criterios técnicos y no políticos» (5).

Sin embargo, y a pesar de las diferencias doctrinales, existe cierto consenso a la hora de analizar la regulación de la IA con la balanza excesivamente inclinada hacia el lado de los riesgos. Particularmente esta distribución de pesos me parece perjudicial cuando tratamos de realizar un análisis jurídico objetivo. Al centrar nuestra atención exclusivamente en conjurar el peligro, desatendemos las oportunidades y beneficios que puede aportarnos la implementación de la IA. En consecuencia, las propuestas tienden a ser restrictivas o excesivamente intervencionistas, lo que implica cargar una enorme losa sobre los hombros de la Administración y los distintos agentes implicados. La virtud, como diría varios siglos atrás un buen amante de la sabiduría, está en el justo medio. Por tanto, las respuestas jurídicas deben ser equilibradas y tener en cuenta todas las aristas del objeto a regular.

(1) Cabe destacar que el Consejo aprobó su propuesta transaccional a la iniciativa de la Comisión el 25 de noviembre de 2022. La propuesta realiza cambios sustanciales tanto en el concepto de IA (más restrictivo) como en las facultades de la Comisión a la hora de adaptar especificaciones comunes. Solo falta que el Parlamento manifiesta su posición para que el procedimiento legislativo siga su curso y el acto legislativo se apruebe definitivamente.

(2) A. BOIX PALOP (2020: 236-238).

(3) A. HUERGO LORA (2020: 64-67).

(4) J. PONCE SOLÉ (2022: 63).

(5) A. CERRILLO I MARTÍNEZ (2020: 24).

En mi opinión, cuando el usuario del sistema sea una Administración Pública debemos partir de una premisa: las tecnologías emergentes deben utilizarse en la medida en que satisfagan los intereses generales y siempre que contribuyan a materializar los principios que han de regir su actuación (6). Si contamos con herramientas que pueden mejorar la eficiencia en la gestión de los recursos públicos y, además, puede incrementarse significativamente la eficacia en la prestación del servicio, ¿no deberíamos esforzarnos más en plantear soluciones en lugar de óbices? ¿Es mejor opción no aprovechar el progreso tecnológico y mantener una Administración caracterizada por su lentitud y un burocratismo elefantásico? En este sentido, la implementación de un modelo de gestión pública basado en la gobernanza inteligente debería ser uno de los principales objetivos a lograr por los poderes públicos (7).

La narrativa apocalíptica nos lleva a un callejón sin salida y no atiende al principio de realidad freudiano. Siguiendo a HUERGO LORA (8) el concepto IA es antropomórfico. Nada tiene que ver la inteligencia humana con la artificial (9). Lo siento por algunos, pero los robots no van a acabar con la humanidad a menos que un ser humano apriete el botón, pero esa es otra historia. En efecto, el peligro está en la persona, no en la máquina. Dejemos los escenarios distópicos a guionistas y escritores de ciencia ficción. Acertadamente, señala BARRIO ANDRÉS que «... tenemos que desmitificar y juridificar esta tecnología [...] A mi juicio, el principal reto de nuestro tiempo [...] es cómo controlar el proceso de toma de decisiones de los sistemas inteligentes...» (10). Abordémosla, pues, como lo que es, una herramienta con un increíble potencial para mejorar la actuación administrativa.

En este sentido, la situación requiere utilizar técnicas jurídicas que nos permitan integrar las nuevas tecnologías en la Administración. ¿Límites? Únicamente aquellos que se traduzcan en garantías para el administrado. Haciendo mío el razonamiento de BOIX PALOP, con el que solo puedo estar de acuerdo en lo relativo a la defensa de la protección jurídica del ciudadano, «... los

(6) El art. 103.1 de nuestra Norma Fundamental es taxativo cuando dispone que «La Administración Pública sirve con objetividad los intereses generales y actúa de acuerdo con los principios de eficacia, jerarquía, descentralización, desconcentración y coordinación, con sometimiento pleno a la ley y al Derecho».

(7) A. CERRILLO I MARTÍNEZ (2019: 6-9).

(8) A. HUERGO LORA (2022: 81).

(9) La confusión que algunas personas sufren cuando interactúan con un sistema informático se denomina «efecto ELIZA». El fenómeno se traduce en que el sujeto afectado cree que la máquina le comprende y es empática, pudiendo llegar a entablar un vínculo estrecho basado en la confianza hacia el sistema. Este es uno de los motivos por los que la propuesta de Reglamento obligará a que en todo momento se advierta al usuario de que está interactuando con una máquina.

(10) M. BARRIO ANDRÉS (2022: 21).

ciudadanos deben conservar todos sus derechos frente a la actuación administrativa también cuando esta es realizada por medios algorítmicos [...] El equilibrio entre la protección jurídica de su posición jurídica y la capacidad de actuación del Estado, del poder, no ha de verse alterada [...] por el simple hecho de que los poderes públicos dispongan de herramientas técnicas que les permitan ser más eficientes en su actuación» (11).

Esta es la razón principal que me impulsa a escribir el presente artículo científico. En las siguientes páginas se pretende aportar una posible solución a la situación de desprotección del administrado frente al empleo de sistemas de IA o automáticos en el procedimiento administrativo. Ello se hará a través del principio que he denominado «presunción de falibilidad». Dicho principio (que opera en aquellos supuestos de decisiones automatizadas desfavorables) podría constituir una salvaguarda de los derechos e intereses legítimos del ciudadano, además de servir de catalizador para la implementación de la IA en procedimientos donde no sea necesario aplicar discrecionalidad o interpretar conceptos jurídicos indeterminados.

La actividad administrativa mediante algoritmos es ya una realidad que debemos regular. Por este motivo, la Academia tiene el deber de construir puentes que faciliten su aplicación en favor del ciudadano, no murallas que frenen el uso de una tecnología que puede contribuir a la satisfacción de los intereses generales.

El objetivo del trabajo es dar una respuesta jurídica al problema actual fruto de la desactualización de nuestro ordenamiento jurídico-administrativo e invitar a la reflexión. En el primer epígrafe, se expondrán los conceptos básicos del aprendizaje automático y las diferencias entre el supervisado y no supervisado. Clarificar estos conceptos nos permitirá facilitar la comprensión del estudio y la hipótesis planteada.

A continuación, se desarrollará el principio de presunción de falibilidad de las decisiones desfavorables dictadas al amparo de una actuación automatizada. En primer lugar, se estudiarán las fuentes de las que se nutre esta garantía jurídica para, posteriormente, analizar sus efectos, así como su ámbito de aplicación. Asimismo, se utilizará el conflicto jurídico derivado del uso del programa informático BOSCO (acrónimo que tiene su origen en la fusión de las palabras bono social comercializadoras) para justificar la necesidad de reconocer este principio en nuestro ordenamiento jurídico-administrativo. Por último, se dedicará el epígrafe final a transmitir unas breves reflexiones conclusivas fruto del análisis realizado.

(11) A. BOIX PALOP (2022: 92-93).

1. Introducción al concepto de aprendizaje automático supervisado y no supervisado (12)

Previamente, y con el ánimo de no ahogarnos en un mar de conceptos ajenos a nuestra disciplina, conviene explicar de forma sencilla qué es un algoritmo, qué se entiende por aprendizaje automático y, dentro de este último, cuál es la diferencia básica entre los modelos supervisados y no supervisados.

A la pregunta, ¿qué es un algoritmo? (13) Básicamente es el conjunto de instrucciones que el programador da al sistema para conseguir un resultado (14). El algoritmo es el *software* en su conjunto. En realidad, este concepto es muy genérico y nos sirve de poco a la hora de abordar el estudio del aprendizaje automático.

Una vez visto el significado del concepto algoritmo, ¿qué se entiende por aprendizaje automático? Aquí la cosa se complica. El aprendizaje automático, en palabras de MITCHELL, «se ocupa de la cuestión de cómo construir programas informáticos que mejoren automáticamente con la experiencia» (15). Un concepto más retórico lo aporta DOMINGOS al manifestar que «El aprendizaje automático es algo nuevo bajo el sol: una tecnología que se construye a sí misma [...] Un algoritmo de aprendizaje es como un maestro artesano, cada una de sus producciones es diferente y está exquisitamente adaptada a las necesidades del cliente» (16). Por su parte, *Google Cloud* lo define como «un subconjunto de inteligencia artificial que permite que un sistema aprenda y mejore de forma autónoma mediante redes neuronales y aprendizaje profundo, sin tener que ser programado explícitamente, a través de la ingesta de grandes cantidades de datos» (17).

Finalmente, ¿qué diferencias existen entre los sistemas de aprendizaje automático supervisados y no supervisados? En la correcta distinción de la técnica está la clave que nos permitirá realizar un análisis jurídico riguroso. Algunos autores distinguen entre «algoritmos predictivos y no predictivos» (18). A pesar de estar de acuerdo con la sistematización realizada por HUERGO LORA,

(12) El estudio se centra en esta rama de la inteligencia artificial porque es la que previsiblemente será objeto de regulación en el futuro proyecto de Reglamento.

(13) Para una definición más desarrollada del concepto, *vid.* A. HUERGO LORA (2020: 26-30).

(14) Siendo precisos, la Real Academia Española define un algoritmo como el «Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema».

(15) T. MITCHELL (1997: XV).

(16) P. DOMINGOS (2015: 11).

(17) <https://cloud.google.com/learn/what-is-machine-learning?hl=es#section-2> (última consulta realizada el 18 de abril de 2023).

(18) A. HUERGO LORA, 2020 (68-87).

en el campo del aprendizaje automático no creo que esta sea la principal diferencia, pues ambos son modelos cuyos resultados son formalmente predicciones basadas en el análisis de datos.

La diferencia principal entre un modelo supervisado y uno no supervisado, teniendo en cuenta lo establecido por los expertos en computación, a mi juicio, es la siguiente: mientras en el primero los datos de entrada (*inputs*) están etiquetados, configurándose el sistema de forma tal que se predetermina la variable de respuesta (*output* o salida); en el segundo, en cambio, no existe una variable de respuesta, sino únicamente los datos de entrada sin etiquetar. Es decir, no se entrena al sistema para que llegue a un resultado determinado. No obstante, ambos sistemas tienen la capacidad de aprender de forma autónoma y predicen resultados.

2. La conveniencia de la implementación de sistemas de aprendizaje automático supervisados en actuaciones que impliquen únicamente el ejercicio de potestades regladas

Como se ha señalado *ut supra*, en los sistemas de aprendizaje automático supervisados es un humano quien predetermina el resultado. A modo de ejemplo, si en un procedimiento administrativo de concesión de subvenciones el sujeto X debe cumplir con los requisitos 1, 2 y 3 para poder optar a la subvención (los requisitos serían las variables explicativas o, en otras palabras, el conjunto de datos etiquetados), en caso de cumplirlos, el propio sistema le dirigirá de forma automática a la variable de respuesta que implique la concesión. En caso de no cumplir las condiciones, el resultado conllevará la denegación. Estos resultados han sido programados previamente en la fase de desarrollo y entrenamiento del modelo, por lo que conocemos cuál será la respuesta en todo momento.

Debido a su naturaleza, el sistema es perfecto para la toma de decisiones que no impliquen discrecionalidad administrativa ni aplicación de conceptos jurídicos indeterminados. Asimismo, este modelo mejora su eficacia con el tiempo, aprende por sí mismo (si se mantienen los datos actualizados) y es fácilmente revisable, ya que el usuario supervisa y predetermina la salida. Si hay desvíos, se corrigen. La ventaja, además de la velocidad en la clasificación de grandes cantidades de información (19), es que el sistema mejora a partir de la experiencia, llegando a un punto en el que, sin necesidad de suministrarle los datos etiquetados, el propio modelo los clasificará y resolverá el problema.

(19) Esta no es la ventaja principal del aprendizaje automático, pues ya existen aplicaciones informáticas que no emplean inteligencia artificial y realizan funciones de clasificación.

Visto el funcionamiento de los modelos supervisados, ¿cómo puede la Administración garantizar los derechos del administrado en el procedimiento cuando se utilizan sistemas de IA supervisados? (20) Aquí entra en juego el principio de presunción de falibilidad que desarrollaremos más adelante. Por ahora solo veremos su ámbito de aplicación, diferenciando entre si el acto es favorable o desfavorable.

En el supuesto de que se trate de un acto favorable, al no requerir la motivación exigida por el art. 35 de la Ley 39/2015, de 1 de octubre, de Procedimiento Administrativo Común de las Administraciones Públicas (en adelante, LPAC) automáticamente se dictará y notificará al interesado (21). Al no existir un perjuicio, el acto administrativo producirá todos sus efectos de acuerdo a lo dispuesto en los arts. 34 y 39 LPAC, lo que permite agilizar los trámites procedimentales. En la esfera de los actos favorables no tiene sentido comprobar todos los resultados dados por el sistema. En este sentido, se presume la fiabilidad del modelo y la supervisión humana queda excluida.

Por el contrario, si nos encontramos ante un acto administrativo desfavorable se dictará un acto de trámite traducido al lenguaje natural y motivado (aquí la explicabilidad es muy sencilla, ya que no existe discrecionalidad administrativa y el error del algoritmo es fácilmente identificable). Tanto el derecho a la buena administración del art. 41 de la Carta de Derechos Fundamentales de la Unión Europea (en adelante, CDFUE) como el art. 35.1.a) LPAC exigen que al ciudadano se le dé una respuesta motivada en aquellos supuestos de actos que afecten a su esfera personal o patrimonial.

Asimismo, el art. 41, apartado 2, *in fine*, de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (en adelante, LRJSP) dispone que en los supuestos de actuación administrativa automatizada «se indicará el órgano que debe ser considerado responsable a efectos de impugnación». Implícitamente podemos extraer del inciso una garantía en favor del administrado, cual es el derecho a impugnar en vía administrativa todo acto desfavorable que haya sido dictado sin intervención humana. La automatización no puede convertirse en una suerte de escudo para el órgano competente que le exima de responsabilidad. Por otra parte, este precepto, desde mi punto de vista, debe interpretarse en el sentido del reconocimiento legal de un deber de responsabilidad de la Administración dueña del programa, no

(20) El razonamiento también se extiende a las actuaciones automatizadas que no emplean IA. Sobre la diferenciación entre sistemas automáticos y autónomos, *vid.* A. SORIANO ARNANZ (2021), «Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos», *Revista de Derecho Público: Teoría y Método*, 3, pág. 88.

(21) Sobre la cuestión, *vid.* E. GAMERO CASADO (2021), «Necesidad de motivación e invalidez de los actos administrativos sustentados en inteligencia artificial o algoritmos», *Almacén de Derecho*, 4.

pudiendo esta derivar la impugnación a órganos administrativos ajenos a la aplicación informática (22).

En conclusión, con esta forma de proceder conseguimos reducir los tiempos en la tramitación y beneficiar a todos aquellos ciudadanos que cumplen los requisitos, a la vez que respetamos las garantías del procedimiento en el caso de actuaciones desfavorables. La Administración humana sólo intervendrá cuando los administrados sufran un perjuicio, ya sea por la utilización de un sistema basado en IA o por la simple actuación automatizada.

3. La problemática del uso de sistemas de inteligencia artificial que utilicen técnicas de aprendizaje automático no supervisado por la Administración. ¿Prohibición absoluta o relativa?

A diferencia de los modelos de IA supervisados, en los sistemas de IA basados en aprendizaje automático no supervisado no se predetermina la salida ni se etiqueta el conjunto de datos suministrado. En consecuencia, será el propio modelo quien busque correlaciones entre los datos aportados y determine la salida. Siguiendo la explicación de SIMEONE, «... el conjunto de entrenamiento consiste en entradas no etiquetadas, es decir, de entradas sin ninguna salida predeterminada» (23). El objetivo general del aprendizaje no supervisado es descubrir propiedades ocultas en los datos. Al modelo no se le entrena en ningún momento para que llegue a un resultado concreto.

Como puede inferirse de la definición, en esta clase de modelos no existe intervención humana que determine el resultado (en nuestro campo, la actuación administrativa algorítmica). Pero, ¿puede explicarse el motivo por el que llega a la decisión la máquina? Sí, aunque en determinados supuestos puede ser sumamente complejo debido a la opacidad del sistema. En este sentido, concluye SIMEONE que «La máquina entrenada es una caja negra que asigna las entradas a las salidas. Como tal, no proporciona medios directos para averiguar por qué se ha producido una salida determinada en respuesta a una entrada, aunque investigaciones recientes han hecho algunos progresos en este frente» (24).

Además, las técnicas empleadas en el aprendizaje no supervisado suelen ser abstractas y discriminatorias, ya que no tienen en cuenta la individualidad

(22) Así ha sucedido con el programa informático BOSCO utilizado por el Ministerio para la Transición Ecológica y el Reto Demográfico para la concesión del bono social eléctrico. Ante la denegación del bono, el ciudadano debe acudir a las autoridades de consumo autonómicas, lo que implica derivar la responsabilidad a una Administración que desconoce el funcionamiento de la aplicación.

(23) O. SIMEONE (2018: 2).

(24) O. SIMEONE (2018: 3).

del sujeto, sino su proximidad a un grupo con el que comparte determinadas características (25). Un ejemplo de su funcionamiento opaco lo encontramos en el programa «SyRI» (26), empleado por la Inspección de Trabajo y Seguridad Social holandesa para detectar y prevenir el fraude en la obtención de prestaciones de la seguridad social. La normativa que regulaba su uso fue declarada contraria al art. 8 del Convenio Europeo de Derechos Humanos, siendo uno de los principales motivos la falta de transparencia del sistema (27).

La alta probabilidad de que existan elementos discriminatorios en la fórmula por la que se llega al resultado unido a la complejidad de explicar su funcionamiento hace recomendable limitar su empleo por las Administraciones Públicas. En esta esfera debe entrar en juego el principio de reserva de humanidad propugnado por PONCE SOLÉ (28).

Dicho esto, ¿en qué ámbitos de actuación sería adecuado utilizar modelos algorítmicos no supervisados? A mi juicio, sería positiva su utilización como herramienta de apoyo para la evaluación de políticas públicas y, por qué no, para la toma de decisiones que queden al margen de procedimientos que impliquen el reconocimiento o denegación de derechos (29). A modo de ejemplo, piénsese en un Consejo de Gobierno autonómico que quiera conocer con mayor precisión dónde sería más adecuado dirigir los recursos disponibles de cara a elaborar las cuentas públicas. Gracias a esta tecnología la política presupuestaria puede verse mejorada.

Por el contrario, aquellos sistemas de IA utilizados para la prevención y detección del fraude (por ejemplo, los algoritmos empleados por la Agencia Tributaria) deberían quedar, de momento, excluidos de plano, pues su funcionamiento es opaco y existe un riesgo alto de que empleen técnicas discriminatorias, contribuyendo, en consecuencia, a la estigmatización de determinados sectores de la población. Idéntica exclusión debe operar cuando exista discrecionalidad en el reconocimiento de derechos o en el acceso a servicios. En el ámbito de la seguridad pública, la primera cámara del Tribunal Constitucional

(25) A. SORIANO ARNANZ, (2021:91-93).

(26) Para profundizar en este asunto, *vid.* L. COTINO HUESO «SyRI, ¿a quién sanciono? Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020», *La Ley Privacidad*, 4, 2020.

(27) Sentencia de 5 de febrero de 2020 del Tribunal de Distrito de la Haya (C / 09/550982 / HA ZA 18-388, ECLI:NL:RBDHA:2020:1878). A través del siguiente enlace puede accederse directamente a su contenido *Rechtbank Den Haag 5 februari 2020*, ECLI:NL:RBDHA:2020:1878.

(28) Sobre el principio de reserva de humanidad, *vid.* J. PONCE SOLÉ (2019), «Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico», *Revista General de Derecho administrativo*, 50.

(29) A. HUERGO LORA (2022: 85).

Federal de Alemania ha prohibido recientemente el uso de sistemas de aprendizaje automático en una sentencia que declara la inconstitucionalidad de dos leyes que permitan a la policía utilizar técnicas basadas en el análisis masivo de datos en tareas de prevención y persecución de delitos. En el párrafo 121 de la sentencia el Alto Tribunal alemán concluye que el uso de estos modelos debe quedar expresamente excluido en la ley (30).

En definitiva, sería adecuado que el Legislador regulara una prohibición relativa de los sistemas algorítmicos que utilicen técnicas de aprendizaje automático no supervisado. Así, mientras se prohíbe su uso en procedimientos que requieran el ejercicio de discrecionalidad o afecten negativamente a los derechos de los administrados, debe autorizarse su empleo como instrumento de evaluación de políticas públicas, así como cuando constituya un apoyo en la toma de decisiones que obedezcan a criterios exclusivamente políticos. Hasta que este tipo de sistemas no sean 100% explicables no pueden ser utilizados por la Administración de forma generalizada.

II. EL PRINCIPIO DE PRESUNCIÓN DE FALIBILIDAD DE LAS DECISIONES DESFAVORABLES DICTADAS POR SISTEMAS ALGORÍTMICOS

El principio de presunción de falibilidad surge a raíz de profundizar en el estudio de la doctrina y partiendo de un hecho que no admite prueba en contrario: la dificultad actual de explicar y comprender el funcionamiento de un sistema algorítmico que opere con técnicas de aprendizaje automático. Teniendo en cuenta que en breve contaremos con un acto legislativo nuevo enfoque que someterá a control previo los sistemas calificados como de alto riesgo antes de su introducción en el mercado (31), el debate jurídico ya no girará en torno a la seguridad del *software*. Así, a partir de la de entrada en vigor de la norma europea los modelos de alto riesgo deberán pasar una evaluación de la conformidad *ex ante* a fin de garantizar que cumplen con los requisitos previstos en aquella.

(30) Un breve comentario a esta sentencia en español lo encontramos en L. COTINO HUESO (2023), «Los requisitos del Tribunal Constitucional alemán para el análisis automatizado de datos o con inteligencia artificial, que no se cumplen en España», blog de la *Revista Catalana de Dret Públic*. Puede accederse al mismo a través del siguiente enlace <https://eapc-rcdp.blog.gencat.cat/2023/04/12/los-requisitos-del-tribunal-constitucional-aleman-para-el-analisis-automatizado-de-datos-o-con-inteligencia-artificial-que-no-se-cumplen-en-espana-lorenzo-cotino-hueso/> (última consulta realizada el 20 de abril de 2023).

(31) Para comprender mejor la futura regulación es necesario contar con un conocimiento básico de las normas técnicas armonizadas. En este sentido, *vid.* V. ÁLVAREZ GARCÍA (2020), *Las normas técnicas armonizadas (Una peculiar fuente del Derecho europeo)*, 1ª ed., Madrid, lustel.

Serán, pues, los organismos notificados y las Autoridades nacionales de supervisión quienes velen por que los proveedores y usuarios cumplen con las prescripciones esenciales del acto legislativo europeo (32). ¿Dónde debería poner el foco el administrativista a partir de la entrada en vigor del Reglamento de IA? En mi opinión, el estudio de la adaptación de los procedimientos administrativos a través de nuevas fórmulas que permitan integrar la IA en el día a día de la Administración.

Al igual que el principio de presunción de legalidad de los actos administrativos constituye un baluarte sobre el cual la Administración ejerce sus prerrogativas exorbitantes para cumplir con su finalidad, la presunción de falibilidad se erigiría como un escudo frente a las decisiones dictadas por sistemas autónomos o automáticos. Su *ratio essendi* es, por tanto, ser garantía jurídica de los derechos e intereses legítimos del administrado, así como límite para los poderes públicos. En este sentido, la lucha contra las inmunidades del poder defendida por GARCÍA DE ENTERRÍA (33) ha de tenerse más presente que nunca. Ningún manto de oscuridad puede arropar la actuación de la Administración.

Dicho lo cual, considero que el ciudadano no necesita acceder al código fuente del programa, por muy loables que sean algunos razonamientos desde el punto de vista de la transparencia. Tampoco lo exige la futura Propuesta de Reglamento (34). Es más, en lo que atañe a la publicidad del código fuente, el futuro Reglamento de IA se decanta por la protección de la propiedad intelectual del sistema (35).

(32) Un estudio sobre la normalización industrial recomendable a todo aquel que quiera profundizar en la regulación de la IA lo encontramos en V. ÁLVAREZ GARCÍA (1999), *La normalización industrial*, 1º ed., Valencia, Tirant lo Blanch.

(33) Vid. E. GARCÍA DE ENTERRÍA (1962), «La lucha contra las inmunidades del poder en el derecho administrativo (poderes discrecionales, poderes de gobierno, poderes normativos)», *Revista de Administración Pública*, 38, págs. 159-208.

(34) El art. 64.2 del AAI permite acceder al código fuente únicamente a las autoridades de vigilancia del mercado. Así, establece el precepto que «En caso necesario y previa solicitud motivada, se concederá a las autoridades de vigilancia del mercado acceso al código fuente del sistema de IA para evaluar la conformidad del sistema de IA de alto riesgo con los requisitos establecidos en el título III, capítulo 2».

(35) El art. 70.1.a) del AAI dispone de forma taxativa que «Las autoridades nacionales competentes y los organismos notificados involucrados en la aplicación del presente Reglamento respetarán la confidencialidad de la información y los datos obtenidos en el ejercicio de sus funciones y actividades de modo que se protejan, en particular:

a) los derechos de propiedad intelectual y la información empresarial confidencial o los secretos comerciales de una persona física o jurídica, *incluido el código fuente*, salvo en los casos contemplados en el artículo 5 de la Directiva 2016/943 relativa a la protección de los conocimientos técnicos y la información empresarial no divulgados (secretos comerciales) contra su obtención, utilización y revelación ilícitas».

Ahora bien, lo que sí necesita el administrado es la existencia de un acto que le permita conocer las razones fácticas y jurídicas de una actuación administrativa que le sea desfavorable. Siguiendo a PONCE SOLÉ, para construir el principio distinguiremos entre dos técnicas reguladoras: la supervisión humana y la reserva de humanidad (36). Desde esta perspectiva, la presunción de falibilidad de los modelos algorítmicos que impliquen decisiones desfavorables para el administrado supondría una forma de hacer efectiva la supervisión humana en el procedimiento administrativo.

Por otra parte, debe recordarse al lector que en este estudio se parte de una hipótesis en virtud de la cual los sistemas de IA basados en aprendizaje automático no supervisado no pueden (o no deben) ser utilizados por las Administraciones Públicas para la toma de decisiones discrecionales. Por ahora, y hasta que no se pueda garantizar la explicabilidad de los resultados al 100%, estos sistemas no deberían utilizarse como fundamento de las decisiones que afecten a derechos o intereses legítimos de los ciudadanos. Ahora bien, el argumento de la empatía esgrimido para justificar la reserva de humanidad carece de fundamento desde una perspectiva iuspositivista, pues, ¿existe en nuestro ordenamiento jurídico-administrativo un derecho del administrado a que las resoluciones se dicten empáticamente? Al funcionario no se le exige que sea empático, sino que desempeñe sus funciones con estricto respeto a la Ley y al Derecho, no a la moral o la ética. De la empatía, a mi juicio, no puede surgir la necesidad de una reserva de humanidad.

En cambio, la Administración sí tiene la obligación de motivar sus actos desfavorables so pena de incurrir en arbitrariedad y, no en base a los resultados probabilísticos dictados por un programa informático, sino con fundamento en hechos y derecho. Por este motivo considero que hoy por hoy es jurídicamente inviable la aplicación del aprendizaje automático no supervisado en el procedimiento administrativo. Como buenamente señala SÁNCHEZ MORÓN, «la motivación es, ante todo, una manifestación y una exigencia del principio de transparencia y del derecho de los ciudadanos a la buena administración» (37).

Pero, yendo más allá, ni aun cumpliéndose los requisitos expuestos por COTINO HUESO de transparencia externa y explicabilidad (38) podría dictarse un acto administrativo que no esté fundado en derecho. El «principio de juridicidad» (39) es el primer muro contra el que choca la idea de implementar

(36) J. PONCE SOLÉ (2022: 63-66).

(37) M. SÁNCHEZ MORÓN (2020: 547).

(38) L. COTINO HUESO (2022: 25 y ss.).

(39) Sobre el principio de legalidad y juridicidad, *vid.* S. MUÑOZ MACHADO (2015), «*Tratado de Derecho Administrativo y Derecho Público General*. T. III: *Los principios de constitucionalidad y legalidad*», 4ª ed., Madrid, Boletín Oficial del Estado, págs. 99-136.

la IA de forma generalizada en la Administración. La actuación administrativa debe de ser fiscalizable conforme a parámetros jurídico-constitucionales, no informáticos. Podemos discutir sobre el tipo de vinculación (negativa o positiva) de la Administración con la Ley y su margen de actuación. Ahora bien, lo que no podemos plantear es un modelo de actuación de las Administraciones Públicas regulado a través de fórmulas ajenas al Derecho.

Un algoritmo basado en cálculos matemáticos no podrá jamás sustituir a una disposición general en su función normativa. De hacer eso, estaríamos socavando los principios estructurales que configuran el Estado de Derecho. Aceptar la naturaleza jurídica de los algoritmos aprendices nos llevaría a sustituir el imperio de la ley por el reino de la arbitrariedad probabilística. Las matemáticas, en definitiva, no pueden sustituir en modo alguno al Derecho en su función ordenadora de la realidad social. Ello me lleva a la necesidad de rechazar la tesis que propone la naturaleza reglamentaria de los algoritmos. En este sentido, a pesar de que BOIX PALOP defiende con vehemencia la naturaleza normativa de algunos programas informáticos, él mismo reconoce que (la cursiva es nuestra) «... esos algoritmos y programas son *empleados como elementos que ayudan* a determinar la concurrencia o no de ciertas circunstancias de hecho o que establecen la conveniencia o no de asociar ciertas consecuencias jurídicas a los hechos disponibles» (40).

A lo largo de las páginas el autor deja claro de forma indirecta el papel de intérpretes que representan los sistemas de IA. Así, señala que «Parafraseando a Lessig, que lo expresaba de una forma levemente diferente, podría decirse que el código fuente con el que se escriben los algoritmos que componen los programas que *evalúan cada vez más circunstancias y deciden cada vez más consecuencias jurídicas* que conviene asociar a cada hecho, al formar parte de las *herramientas que emplea el Derecho* para dar respuesta a y ordenar la realidad social...» (41).

Los programas *ayudan y evalúan, sí, pero* tanto las circunstancias fácticas como las consecuencias jurídicas deben estar previstas en una norma jurídica (no surgen del sistema), cuestión que el autor pasa por alto y, a mi parecer, es un aspecto dogmático sustancial a tener en cuenta si la intención es construir una teoría que defienda la naturaleza reglamentaria de los algoritmos. En términos estrictamente técnicos, el código fuente no innova el ordenamiento jurídico-administrativo. Ergo, no es fuente del Derecho ni se integra en el sistema normativo. Que sea un modelo de inteligencia artificial en lugar de un funcionario el que aprecie si concurren los supuestos de hecho y las consecuencias jurídicas preestablecidas en una disposición normativa, no puede

(40) A. BOIX PALOP (2020: 234).

(41) A. BOIX PALOP (2020: 235).

llevarnos a afirmar que los «algoritmos son reglamentos». Si siguiéramos este razonamiento, terminaríamos concluyendo que todos los operadores jurídicos que realizan funciones interpretativas son parte también del ordenamiento jurídico porque evalúan circunstancias fácticas y aplican consecuencias jurídicas.

En mi opinión, todo lo más que se produce es una subdelegación de la función interpretativa (que ya es bastante), siendo esta transferida del funcionario a la máquina por su mayor capacidad de procesamiento; pero en ningún caso nos encontramos ante un ejercicio de la potestad normativa (42). La conclusión a la que llega la teoría no es válida porque, a mi juicio, el autor equipara la función interpretativa con la normativa. Quizás este razonamiento pueda tener acogida en sistemas anglosajones (mucho más flexibles en lo que respecta al sistema de fuentes), pero en sistemas continentales como el nuestro no tiene mucho recorrido. Ahora bien, cuestión distinta es que el Legislador opte por aplicar el procedimiento para la elaboración de Reglamentos de la LPAC a los sistemas de inteligencia artificial empleados por la Administración para dotar a la herramienta de cierta transparencia y legitimidad democrática. Pero para eso no es necesario calificar a los algoritmos como normas jurídicas. Ello sería una mera *fictio iuris* innecesaria, pues huelga decir que toda actuación de la administración (sea a través de algoritmos o por otros medios) está sometida a la Constitución, a la ley, y a los principios generales del Derecho que informan nuestro ordenamiento.

En fin, a pesar de las discrepancias teóricas, comparto plenamente la defensa de la primacía del Derecho realizada por el autor cuando afirma «la necesaria preeminencia jurídica tanto para la predeterminación típica requerida para la aplicación posterior al caso concreto de las reglas también cuando se haga por medio de estos nuevos instrumentos que, junto con las normas jurídicas tradicionales, conforman como parte esencial el nuevo entramado normativo: los algoritmos y los programas y modelos que estos componen y el código fuente en que estos están escritos» (43).

(42) Algo que, por ejemplo, sí ocurre en el ámbito de las normas técnicas armonizadas, consideradas jurisprudencialmente como actos jurídicos de la Unión. A pesar de su naturaleza, estas normas (elaboradas por los organismos de normalización) no cumplen formalmente con el requisito de publicidad normativa (sólo se publican sus referencias). A nivel nacional, nuestro ordenamiento jurídico también contiene disposiciones reglamentarias que utilizan la técnica del reenvío normativo a normas técnicas voluntarias, las cuales, debido a la remisión, se vuelven vinculantes, y ello sin haber sido objeto de publicación. Sobre la cuestión, *vid.* V. ÁLVAREZ GARCÍA (2022), «La problemática de la publicidad oficial de las normas técnicas de origen privado que despliegan efectos jurídico-públicos», *Revista de Derecho Comunitario Europeo*, 72, págs. 449-482; sobre la capacidad normativa de los sujetos privados y la subdelegación de la potestad reglamentaria, *vid.* V. ÁLVAREZ GARCÍA (2010), *Industria*, 1ª ed., Madrid, Lustel, págs. 224 y ss.

(43) A. BOIX PALOP (2020: 233).

1. Su conexión con el derecho a una buena administración y la Carta de Derechos Digitales del Gobierno de España

En cuanto a las fuentes (normativas y no normativas) de las que bebe este principio, podemos diferenciar dos: el derecho a la buena administración previsto en la CDFUE (44) y la Carta de Derechos Digitales aprobada por el Gobierno de España en el año 2021. También ha de tenerse en cuenta lo dispuesto en el art. 41 de la Ley 40/2015, de 1 de octubre, del Régimen Jurídico del Sector Público (en adelante, LRJSP).

El art. 41.2 a) de la CDFUE reconoce «el derecho de toda persona a ser oída antes de que se tome en contra suya una medida individual que la afecte desfavorablemente». Como puede observarse, la Carta prevé un derecho de audiencia en caso de actuaciones desfavorables para el interesado. Junto al derecho a ser oído, dispone el apartado c) del mismo precepto «la obligación que incumbe a la administración de motivar sus decisiones». Respecto a la obligación de motivación reconocida por en la CDFUE, señala Sánchez Morón que «...la exigencia legal de motivación no se ciñe a aquellos casos en que el acto afecte desfavorablemente a derechos e intereses individuales legítimos, sino que alcanza a otros muchos, en el derecho europeo prácticamente a todos los actos jurídicos» (45).

Aplicado al ámbito de las decisiones algorítmicas, todo ciudadano tiene derecho a conocer las razones fácticas y jurídicas (que no informáticas) del acto administrativo dictado. El principio de presunción de falibilidad quedaría configurado como una garantía del administrado en aquellos supuestos de decisiones desfavorables cuyo fundamento resida en los resultados de un sistema algorítmico. Gracias a su aplicación un acto administrativo desfavorable dictado de forma automática no podría producir efectos *ipso facto*. La presunción, por tanto, hace decaer la ejecutividad del acto y obliga a que un humano revise y motive la decisión, dándose audiencia al interesado.

Por otra parte, y aunque carezca de valor normativo, es importante destacar el artículo XXV, apartado 3, letra b), el cual dispone que «Las personas tienen derecho a solicitar una supervisión e intervención humana y a impugnar las decisiones automatizadas tomadas por sistemas de inteligencia artificial que produzcan efectos en su esfera personal y patrimonial». El precepto es algo genérico en su redacción, aunque considero que ha de interpretarse en el sentido de que el derecho a la supervisión debe surgir cuando existan efectos desfavorables.

(44) Para un estudio específico sobre el particular, *vid.* J. VALERO TORRUIOS (2019), «Les garanties jurídiques de la intelligència artificial en l'activitat administrativa des de la perspectiva de la bona administració», *Revista Catalana de Dret Públic*, 58, págs. 82-96.

(45) M. SÁNCHEZ MORÓN (2020: 547).

A continuación, en el apartado 2 del artículo XXVII (rubricado «Garantía de los derechos en entornos digitales») se señala que «Asimismo, se promoverá la garantía de los derechos reconocidos en esta Carta en el marco de las relaciones con la Administración de Justicia y, particularmente, los derechos relacionados con la inteligencia artificial, cuando se recurra a ésta para la utilización o el desarrollo de sistemas de soporte a las decisiones o de herramientas de justicia predictiva». Como puede observarse, la Carta de Derechos Digitales se desprende la voluntad del Poder Ejecutivo estatal de regular mecanismos de supervisión y revisión de aquellas decisiones automatizadas que afecten a los derechos de los ciudadanos. Luego, una de las garantías procedimentales a tener en cuenta en una futura propuesta normativa podría ser la presunción de falibilidad.

¿Qué se consigue reconociendo este principio procedimental? Impedir que se dicten resoluciones desfavorables basadas únicamente en los resultados de un programa informático. En virtud del mismo, (siempre que se restrinjan derechos o intereses legítimos) se presumirá que la máquina ha fallado y el órgano competente estará obligado a traducir el resultado en un acto administrativo (en este caso, de trámite) que cumpla los requisitos del art. 35 LPAC.

Precisamente el déficit de garantías es el principal motivo por el que miles de ciudadanos ven rechazado su derecho al bono social eléctrico (46). Y ello porque las comercializadoras de referencia fundamentan la denegación de la ayuda en base a los resultados del programa informático BOSCO utilizado por el Ministerio para la Transición Ecológica y el Reto Demográfico. La prueba principal de lo que aquí se expone se encuentra recogida en el Informe anual 2022 del Defensor del Pueblo, donde queda patente el fallo del algoritmo y la falta de explicabilidad de las decisiones automatizadas desfavorables (47).

En su Informe, el Defensor del Pueblo propuso al Ministerio para la Transición Ecológica y el Reto Demográfico «que valorara la posibilidad de que a los solicitantes del bono social eléctrico, a quienes se les haya denegado esta condición, pudieran acudir directamente ante las administraciones competentes en materia de energía. De esta manera, todo el proceso de verificación podría hacerse de una manera más rápida y eficiente» (48). La propuesta fue rechazada.

(46) La regulación de esta ayuda social se encuentra recogida en el art. 45 de la Ley 24/2013, de 26 de diciembre, del Sector Eléctrico, desarrollado mediante el Real Decreto 897/2017, de 6 de octubre, por el que se regula la figura del consumidor vulnerable, el bono social y otras medidas de protección para los consumidores domésticos de energía eléctrica y la Orden ETU/943/2017, de 6 de octubre, por la que se desarrolla el Real Decreto 897/2017, de 6 de octubre, por el que se regula la figura del consumidor vulnerable, el bono social y otras medidas de protección para los consumidores domésticos de energía eléctrica.

(47) Informe Anual del Defensor del Pueblo (2022: 101).

(48) Informe Anual del Defensor del Pueblo (2022: 102).

Actualmente, la única vía para impugnar la denegación del bono social eléctrico es presentar una reclamación ante los servicios de consumo (49). Mientras se tramita el procedimiento, personas vulnerables y con escasos recursos quedan a merced de la decisión desfavorable tomada por un algoritmo, a pesar de cumplir todos los requisitos para obtener el bono. Aplicando el principio de presunción de falibilidad evitaríamos situaciones como esta, pues, antes de dictarse la decisión desfavorable, la Administración comprobaría de oficio el resultado, dictando, una vez verificados los requisitos, el acto administrativo que proceda.

Teniendo en cuenta que los supuestos de vulneración de derechos pueden multiplicarse exponencialmente a medida que se generalice el empleo de sistemas de IA y se avance en la automatización de los procedimientos, la constitución de nuevas garantías jurídicas es obligatoria.

2. La teoría de la proporcionalidad inversa aplicada a la actuación administrativa automatizada: la prohibición de actos desfavorables cuyo único fundamento sea el resultado del algoritmo y su correlativa permisibilidad en el supuesto de actos favorables

Ya hemos visto el funcionamiento de la presunción de falibilidad, así como su fundamento. No obstante, reflexionando sobre su operatividad real podemos llegar a la conclusión siguiente: si aplicamos el principio de forma generalizada volveríamos al punto de partida. Es decir, a la intervención humana total. No tiene mucho sentido. De ahí que debamos acudir a la proporcionalidad inversa (el concepto es un mero símil para explicar la virtualidad de sus efectos).

Como ya se ha adelantado *ut supra*, debido a su naturaleza garantista la presunción sólo debe operar en aquellos supuestos de actos desfavorables. Por ende, cuando el acto sea favorable y suponga la finalización del procedimiento producirá eficacia inmediata, operando en su lugar el principio de fiabilidad del modelo (50). A modo de ejemplo, piénsese en un procedimiento de concesión de subvenciones. Si de mil solicitudes presentadas novecientas son favorables y cien desfavorables, las primeras tendrán la naturaleza de resolución a todos los efectos (sin perjuicio de su carácter provisional, que no impedirá su eficacia).

Las segundas, por el contrario, tendrán la naturaleza de actos de trámite y se presumirán contrarias a derecho por mor de la presunción de falibilidad.

(49) Art. 8.2 del Real Decreto 897/2017, de 6 de octubre, por el que se regula la figura del consumidor vulnerable, el bono social y otras medidas de protección para los consumidores domésticos de energía eléctrica.

(50) A. CERRILLO I MARTÍNEZ (2020: 34-36).

El órgano competente, después de comprobar los resultados del sistema y la solicitud del interesado, podrá:

1. Corregir el fallo del algoritmo y dictar una resolución favorable, o,
2. Confirmar el resultado.

En el último supuesto, el órgano deberá dictar una propuesta con referencia a los hechos y fundamentos de derecho, a fin de que el administrado conozca las razones de la desestimación. Otorgar a las decisiones de los sistemas algorítmicos esta doble naturaleza (acto definitivo si se trata de una decisión favorable y acto de trámite a la inversa) podría ser una solución que permitiera el uso de la IA supervisada con plenas garantías.

Me parece importante insistir en que en este artículo se defiende únicamente el uso de sistemas de IA supervisados para la tramitación de procedimientos que no impliquen discrecionalidad administrativa cuando afecte a derechos e intereses legítimos del interesado. Evidentemente, todos aquellos sistemas donde se desconozcan los parámetros del algoritmo y las variables de respuesta (los resultados) no pueden ni deben ser utilizados por la Administración para resolver esta clase de procedimientos. ¿Por qué? Porque el usuario (la Administración) no cuenta con la información suficiente para justificar la causa de la decisión. En términos jurídicos, no se cumple con la exigencia de motivación (51).

En consecuencia, la opacidad de estos modelos dificulta su utilización en determinadas parcelas de la actuación administrativa. Ni siquiera como elemento de apoyo a la decisión final, pues, en mi humilde opinión, se estaría incurriendo en una palmaria vulneración del art. 9.3 de nuestra Norma Fundamental en su vertiente de interdicción de la arbitrariedad de los poderes públicos. Además, es muy fácil caer en el sesgo de automatización y fiarse de la supuesta objetividad del sistema (52).

3. El programa BOSCO como ejemplo paradigmático de la necesidad de regular la presunción de falibilidad

En primer lugar, conviene explicar qué es el programa BOSCO desde el punto de vista informático. Ya se ha adelantado *ut supra* que es una herra-

(51) Cuando la explicabilidad, transparencia e interpretabilidad de los sistemas sea una realidad, el informe donde se recoja el motivo que ha llevado a la máquina a dictar un resultado concreto podrá adherirse a la resolución. La motivación *in aliunde* no nos es ajena y llegará el día en que la Administración pueda apoyarse en los resultados del modelo para reforzar todo tipo de decisiones. Ahora bien, jamás se podrá prescindir de las razones fácticas y jurídicas.

(52) J. PONCE SOLÉ (2022: 60).

mienta desarrollada por la Administración y utilizada por las comercializadoras de referencia para decidir quién cumple los requisitos para acceder al bono social eléctrico.

Ahora bien, ¿qué tecnología emplea el sistema? A pesar del déficit de información, el modelo es muy probable que se trate de un mero sistema automático, en contraposición a los sistemas autónomos, siendo estos últimos los únicos que emplean técnicas de aprendizaje automático. Siguiendo la distinción realizada por SORIANO ARNANZ, «Los sistemas automáticos serían aquellos en los que el resultado se encuentra previamente determinado y que no disponen de la capacidad de actualizarse ni tomar decisiones de manera autónoma» (53). Quedan, por tanto, fuera del concepto IA pero dentro de la esfera de las actuaciones administrativas automatizadas.

¿Por qué falla la máquina? Pueden existir varios motivos, como, por ejemplo, que el modelo no esté bien ajustado o que le falten variables para determinar el resultado. Uno de los principales fallos denunciados por la Fundación CIVIO era que a las personas viudas con rentas bajas siempre se les denegaba la solicitud, a pesar de cumplir los requisitos (54). En principio, aquí no existe un error de programación, sino que el modelo no da la respuesta correcta porque no cuenta con todas las variables explicativas.

Para muestra, un botón. El Fundamento de Derecho Tercero de la Sentencia del Juzgado Central de lo Contencioso Administrativo (nº 8) 143/2021, de 30 de diciembre (55) refuerza la tesis de la falta de variables explicativas cuando señala la causa del error. Así, concluye el Magistrado que «en los supuestos de “imposibilidad de cálculo”, por no contar con información suficiente del solicitante en las bases de datos conectadas a la aplicación, se le requiere al mismo para que aporte la documentación, y en base a ella se determina si existe el derecho al bono social o no. Son supuestos en que no es posible el cálculo automático mediante la aplicación, pero no supuestos de denegación del bono social».

Por otra parte, resulta interesante pararse en el razonamiento del Magistrado, pues, si bien reconoce la naturaleza de acto administrativo al resultado dictado por la aplicación informática, parece olvidarse de las garantías y requisitos que debe cumplir la actuación administrativa automatizada. ¿Cumple el programa BOSCO el art. 41 de la Ley de Régimen Jurídico del Sector

(53) A. SORIANO ARNANZ (2022: 13-14).

(54) Un ejemplo de la carta remitida por las comercializadoras denegando el acceso al bono social podéis encontrarlo en el siguiente enlace <https://www.documentcloud.org/documents/5805297-Respuesta-Viudedad.html> (última consulta realizada el 17 de abril de 2023).

(55) Sentencia del Juzgado Central de lo Contencioso Administrativo 143/2021, de 30 de diciembre de 2021 (Roj: SAN 5863/2021 — ECLI:ES:AN:2021:5863).

Público? (56) ¿Pasaría el filtro del artículo 22 del Reglamento General de Protección de datos? (57) ¿Es un acto de trámite o definitivo? Es necesario despejar todas las incógnitas para determinar la legalidad de la actuación. La Administración no es libre para actuar como le venga en gana, pues se halla sometida al principio de legalidad.

Salvo opinión mejor fundada, por lo que se refiere a la naturaleza del resultado de la aplicación solo caben dos opciones, entender que nos encontramos ante un acto definitivo o ante un acto de trámite cualificado. El motivo de este razonamiento viene motivado porque en la sentencia, tras reconocerse en el Fundamento de Derecho Tercero que el acto administrativo es dictado por el Ministerio para la Transición Ecológica utilizando la aplicación, señala a continuación que «Siendo lo anterior así, no puede considerarse que el acto administrativo se dicte por una aplicación informática, sino por un órgano administrativo, y en caso de que el destinatario de dicho acto esté disconforme con el mismo, podrá impugnarlo en vía administrativa, y en vía judicial».

Por tanto, a la vista de la normativa que regula el acceso al bono social y las normas de procedimiento, podría concluirse que la Administración General del Estado está incurriendo en un manifiesto incumplimiento de la legalidad.

(56) Establece el apartado «En caso de actuación administrativa automatizada deberá establecerse previamente el órgano u órganos competentes, según los casos, para la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente. Asimismo, se indicará el órgano que debe ser considerado responsable a efectos de impugnación».

(57) Reconoce el precepto que «1. Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar.

2. El apartado 1 no se aplicará si la decisión:

a) es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento;

b) está autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, o

c) se basa en el consentimiento explícito del interesado.

3. En los casos a que se refiere el apartado 2, letras a) y c), el responsable del tratamiento adoptará las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión.

4. Las decisiones a que se refiere el apartado 2 no se basarán en las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, salvo que se aplique el artículo 9, apartado 2, letra a) o g), y se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado».

Para un estudio sobre el art. 22 RGPD en el ámbito de las decisiones automatizadas, vid. A. PALMA ORTIGOSA (2019), «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos», *Revista General de Derecho Administrativo*, 50.

Las garantías previstas en la normativa que regula el bono son mínimas y el acto administrativo adolece de vicios gravísimos de forma y contenido que afectan a su validez.

Respecto al acceso a la información básica del algoritmo, ¿necesitamos acceder al código fuente para identificar la causa de la denegación del bono social? A mi entender, en este caso no es un requisito indispensable. Acceder al código fuente sólo permitirá al interesado conocer la razón técnica del fallo, pero nada más (58). El programa BOSCO aplica el derecho al caso concreto y, en supuestos como el que nos ocupa, el código fuente no es más que la traducción de una norma jurídica a lenguaje informático (59).

Por este motivo me parece irrelevante conocer con exactitud las instrucciones del modelo (60). ¿Dónde radica el vicio de invalidez más importante entonces? Sin duda alguna, en el déficit de garantías procedimentales, no en la falta de transparencia. Los administrados se han visto perjudicados por la decisión de una máquina que no ha aplicado la norma correctamente. Pero, no solo eso, también sin que se les haya dado la oportunidad de ser oídos y, además, sin conocer las razones exactas del sentido del acto administrativo (61).

De ahí que el que suscribe defienda la necesaria legalización de nuevas garantías que permitan a la Administración aprovechar las ventajas del progreso tecnológico sin merma de los derechos del ciudadano. Si hubiera existido el principio de presunción de falibilidad probablemente ni un solo ciudadano hubiera quedado excluido del acceso al bono social eléctrico injustamente. Todas aquellas decisiones desfavorables hubieran sido remitidas a la unidad administrativa u órgano competente para verificar si se cumplen los requisitos y, en consecuencia, dictar el acto motivado que proceda.

En definitiva, los fallos originados por el programa BOSCO unido al déficit de garantías manifiestan la necesidad de reformas estructurales en nuestro

(58) Para una interpretación a favor del acceso al código, *vid.* M. FUERTES (2022), «Reflexiones ante la acelerada automatización de actuaciones administrativas», *Revista Jurídica de Asturias*, 45, pp. 105-124; y G. VESTRI (2022), «El acceso a la información algorítmica a partir del caso bono social vs. Fundación Ciudadana Civio», *Revista General de Derecho Administrativo*, 61.

(59) A. HUERGO LORA (2020: 85).

(60) No tiene sentido. Podría tenerlo en el supuesto de toma de decisiones que impliquen discrecionalidad administrativa, pero no aquí. Desde el principio se conoce si se cumplen o no los requisitos. Acceder al código fuente sirve para conocer con precisión el funcionamiento del modelo y puede ser útil para saber por qué se ha llegado a un resultado concreto en modelos no supervisados donde exista discrecionalidad administrativa o se apliquen conceptos jurídicos indeterminados.

(61) Las comercializadoras de referencia se limitan a trasladar la causa de la imposibilidad de cálculo dada por la aplicación, cuya explicación suele reducirse a la frase «imposibilidad de comprobar los niveles de renta».

procedimiento administrativo. El asunto BOSCO es solo uno de los muchos problemas jurídicos que surgirán en el pedregoso camino de la Administración hacia la efectiva transformación digital. Avanzar en la implementación de nuevas tecnologías exige la adecuación de los marcos normativos, de tal manera que puedan utilizarse respetando y reforzando los derechos del interesado.

III. REFLEXIONES FINALES: SOBRE LA NECESIDAD DE REMODELAR UNA ARQUITECTURA ADMINISTRATIVA ANQUILOSADA

Los resultados de la investigación evidencian la necesidad de actualizar los marcos regulatorios del procedimiento administrativo común con la finalidad de introducir nuevas garantías jurídicas que protejan ciudadano. La actuación administrativa automatizada, ya sea a través de sistemas automáticos o autónomos, es posible, necesaria y recomendable. Las Administraciones Públicas deben aprovechar los frutos del progreso tecnológico, pero, eso sí, cumpliendo con el principio de legalidad y siempre que su empleo sirva a la satisfacción de los intereses generales.

En este sentido, la doctrina administrativista española se halla en un estadio embrionario, por lo que nos encontramos ante un campo de estudio apasionante y en el que todo está por hacer. Actualmente, aparte de las referencias a principios como el de supervisión humana o reserva de humanidad, no contamos con propuestas de reformas normativas desarrolladas. Por ello, aprovecho la ocasión para reivindicar a la Academia un estudio profundo de la cuestión que vaya más allá de las disquisiciones sobre los desafíos, riesgos y oportunidades de la inteligencia artificial.

Con el presente trabajo se pretende contribuir al desarrollo del estado de la cuestión planteando la introducción de una nueva garantía en nuestro ordenamiento jurídico-administrativo. Así, la presunción de falibilidad de las decisiones desfavorables dictadas por sistemas algorítmicos autónomos o automáticos puede constituir una salvaguarda a los derechos e intereses legítimos de los administrados. Este principio garantiza la intervención humana en las actuaciones administrativas desfavorables, y, a su vez, permite la plena automatización sin necesidad de intervención en el supuesto de actos favorables. En el caso de inexistencia de perjuicios, se presume la fiabilidad del sistema.

Por otra parte, se concluye que los sistemas de inteligencia artificial no supervisados no pueden ser utilizados cuando se ejerzan potestades discrecionales o haya que aplicar conceptos jurídicos indeterminados. En consecuencia, el principio de presunción de falibilidad solo operaría cuando se ejerciten potestades regladas que no impliquen la posibilidad de toma de decisiones alternativas o la interpretación de conceptos jurídicos indeterminados. Debe regir aquí, pues, la reserva de humanidad. Empero, no por la falta de empatía

de las máquinas, sino por la dificultad de cumplir con el requisito de motivación y los principios constitucionales de legalidad y de interdicción de la arbitrariedad de los poderes públicos. La empatía no es un derecho del administrado ni se le exige al funcionario que dicta resolución. Tampoco informa con carácter general el ordenamiento jurídico. La «actuación empática» puede servir como fundamento en la esfera de la ética y la moralidad, cuestiones ambas ajenas al Derecho Administrativo y a la dogmática jurídica. Es más, la razón esencial por la que el legislador debe prohibir el uso de modelos de IA en determinados ámbitos es simple: sus razonamientos están basados, principalmente, en lógica probabilística, no en lógica jurídica.

La exigencia de motivación de los actos administrativos derivada de la LPAC, el respeto a los principios generales del derecho, así como el derecho a una buena administración establecido por la CDFUE obligan a las administraciones públicas a explicar las razones fácticas y jurídicas de la decisión cuando afecte a derechos e intereses legítimos. En el caso que nos ocupa y en aras de la flexibilidad, solo se exigiría dicho requisito en el supuesto de actos desfavorables. La explicabilidad de los modelos de IA supervisados o sistemas automáticos es sencilla, pues los resultados están predeterminados. Sin embargo, no ocurre lo mismo con los modelos no supervisados, de ahí que deba restringirse al máximo su uso hasta que pueda garantizarse la interpretabilidad y transparencia del algoritmo.

También se concluye que el art. 41 LRJSP, en lo que respecta a la impugnación de la actuación administrativa automatizada, debería interpretarse en el sentido de que el órgano responsable tiene que ser el mismo que ha dictado el acto impugnado, o al menos, ser un órgano jerárquicamente superior. Esto no ocurre a fecha de hoy con el programa BOSCO, cuyos resultados se han calificado como actos administrativos dictados por el Ministerio para la Transición Ecológica pero solo pueden ser recurridos ante las autoridades autonómicas de consumo.

En definitiva, la regulación europea de la IA unida a un avance tecnológico vertiginoso obligará a los administrativistas a remodelar la arquitectura de la Administración. La cuarta revolución industrial es una realidad que ha venido para transformarlo todo. Por supuesto, en esa transformación deben incluirse las formas en que la Administración se relaciona con el ciudadano, así como el modo en que se prestan y gestionan los servicios públicos. La gobernanza inteligente es el modelo óptimo para la gestión pública en el siglo XXI y es obligación de los poderes públicos avanzar hacia su implementación efectiva.

Ni los peligros ni los riesgos potenciales pueden ser una excusa para frenar el desarrollo de una tecnología cuya capacidad para mejorar la vida de las personas es indudable. El análisis coste-beneficio es claro y petrificar a la Administración en un marco de actuación fruto del siglo XX no es la solución.

El Derecho debe evolucionar a la par que lo hace la sociedad, pues aquel no es más (o no debería ser más) que el fiel reflejo de la comunidad a la que sirve. El Estado y los poderes públicos que lo integran, como instrumentos al servicio de la sociedad civil, tienen la obligación de adaptarse y aprovechar las nuevas tecnologías para maximizar la calidad de vida de los ciudadanos. Y en ese bienestar, no nos quepa duda, debemos incluir la forma en la que se prestan y se accede a los servicios públicos.

¿Cuál será una de las misiones más importantes la misión del iuspublicista en los próximos años? Hallar las fórmulas jurídicas necesarias para lograr que la integración de esta tecnología se haga con pleno respeto a los derechos del administrado y, además, contribuya a reforzarlos. El progreso tecnológico no es algo nuevo para el Derecho Administrativo, por lo que no conviene caer en la tentación de realizar interpretaciones adanistas y alejadas de la realidad mientras relegamos la lógica jurídica al trastero del olvido. Por este motivo, me parece adecuado terminar con una frase del Profesor ESTEVE PARDO recogida por el maestro Sebastián MARTÍN-RETORTILLO hace más de veinte años y que convendría que todos la tuviéramos bien presente: «Es obligado evitar que el desfallecer o el silencio del derecho nos sitúe irremediamente ante el señorío exclusivo de lo técnico» (62).

IV. BIBLIOGRAFÍA

- ÁLVAREZ GARCÍA, Vicente (1999): *La normalización industrial*, 1º ed., Valencia, Universitat de València.
- (2010): *Industria*, 1º ed., Madrid, Iustel.
- (2020): *Las normas técnicas armonizadas: (Una peculiar fuente del Derecho europeo)*, 1º ed., Madrid, Iustel.
- (2022): «La problemática de la publicidad oficial de las normas técnicas de origen privado que despliegan efectos jurídico-públicos», *Revista de Derecho Comunitario Europeo*, 72, pp. 449-482.
- BARRIO ANDRÉS, Moisés (2022): «Inteligencia artificial: origen, concepto, mito y realidad», *El Cronista del Estado Social y Democrático de Derecho*, 100, pp. 14-21.
- BOIX PALOP, Andrés (2020): «Los algoritmos son reglamentos: La necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la Administración para la adopción de decisiones», *Revista de Derecho Público: Teoría y Método*, 1, pp. 223-270.

(62) S. MARTÍN-RETORTILLO BAQUER (2002: 1476).

- (2022): «Transparencia en la utilización de inteligencia artificial por parte de la Administración», *El Cronista del Estado Social y Democrático de Derecho*, 100, pp. 90-105.
- CERRILLO I MARTÍNEZ, Agustí (2019): «El impacto de la inteligencia artificial en el derecho administrativo, ¿nuevos conceptos para nuevas realidades técnicas?», *Revista General de Derecho Administrativo*, 50.
- (2020): «¿Son fiables las decisiones de las Administraciones públicas adoptadas por algoritmos?», *European review of digital administration & law*, 1.1, pp. 18-36.
- COTINO HUESO, Lorenzo (2019): «Riesgos e impactos del Big Data, la inteligencia artificial y la robótica: enfoques, modelos y principios de la respuesta del derecho», *Revista general de Derecho administrativo*, 50, pp. 1-37.
- (2020): «SyRI, ¿a quién sanciono? Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020», *La Ley Privacidad*, 4.
- (2022): «Transparencia y explicabilidad de la inteligencia artificial y “compañía” (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad...). Para qué, para quién y cuánta», en Cotino Hueso Lorenzo y Jorge Castellanos Claramunt (eds.), *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo Blanch, pp. 29-70.
- (2023): «Los requisitos del Tribunal Constitucional alemán para el análisis automatizado de datos o con inteligencia artificial, que no se cumplen en España», blog de la Revista Catalana de Dret Públic.
- DOMINGOS, Pedro (2015): *The master algorithm: How the quest for the ultimate learning machine will remake our world*, Nueva York, Basic Books.
- FUERTES, Mercedes (2022): «Reflexiones ante la acelerada automatización de actuaciones administrativas», *Revista Jurídica de Asturias*, 45, pp. 105-124.
- GAMERO CASADO, Eduardo (2021): «Necesidad de motivación e invalidez de los actos administrativos sustentados en inteligencia artificial o algoritmos», *Almacén de Derecho*, 4.
- GARCÍA DE ENTERRÍA, Eduardo (1962): «La lucha contra las inmunidades del poder en el derecho administrativo (poderes discrecionales, poderes de gobierno, poderes normativos)» *Revista de administración pública*, 38, pp. 159-208.
- HUERGO LORA, Alejandro y DÍAZ GONZÁLEZ, Gustavo Manuel (2020): *La regulación de los algoritmos*, 1ª ed., Aranzadi.
- (2022): «Gobernar con algoritmos, gobernar los algoritmos», *El Cronista del Estado Social y Democrático de Derecho*, 100, pp. 80-89.

- MARTÍN-RETORTILLO BAQUER, Sebastián (2002): «Breves acotaciones sobre el desarrollo tecnológico y el sistema jurídico-administrativo», en *Nuevas perspectivas del régimen local: estudios en homenaje al profesor José M^o Boquera Oliver*, Valencia, Tirant lo Blanch, pp. 1475-1488.
- MITCHELL, Tom (1997): *Machine Learning*, 1^a ed., Nueva York, McGraw-Hill.
- MUÑOZ MACHADO, Santiago (2015): «*Tratado de Derecho Administrativo y Derecho Público General. Tomo III. Los principios de constitucionalidad y legalidad*», Boletín Oficial del Estado, 4^a Edición.
- PALMA ORTIGOSA, Adrián (2019): «Decisiones automatizadas en el RGPD. El uso de algoritmos en el contexto de la protección de datos», *Revista General de Derecho Administrativo*, 50.
- PONCE SOLÉ, Juli (2019): «Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico», *Revista general de Derecho administrativo*, 50.
- (2022): «Reserva de humanidad y supervisión humana de la Inteligencia artificial», *El Cronista del Estado Social y Democrático de Derecho*, 100, pp. 58-67.
- SÁNCHEZ MORÓN, Miguel (2020): «*Derecho administrativo, Parte General*», 17^a ed., Madrid, Tecnos.
- SIMEONE, Osvaldo (2018): «A very brief introduction to machine learning with applications to communication systems», *IEEE Transactions on Cognitive Communications and Networking*, 4.4, págs. 648-664.
- SORIANO ARNANZ, Alba (2021): «Decisiones automatizadas: problemas y soluciones jurídicas. Más allá de la protección de datos», *Revista de derecho público: teoría y método*, 3, pp. 85-127.
- (2022): «El uso de la inteligencia artificial en la selección y gestión del personal de las administraciones públicas», *Documentación Administrativa. Nueva época*, 9, pp. 11-25.
- VALERO TORRIJOS, Julián (2019): «Les garanties jurídiques de la intelligençia artificial en l'activitat administrativa des de la perspectiva de la bona administració», *Revista Catalana de Dret Públic*, 58, pp. 82-96.
- VESTRI, Gabriele (2022): «El acceso a la información algorítmica a partir del caso bono social vs. Fundación Ciudadana Civio», *Revista General de Derecho Administrativo*, 61, pp. 32.